

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 02:51:12 ; Search time 1626.01 Seconds  
(without alignments)  
13446.329 Million cell updates/sec

Title: US-09-768-781-1

Perfect score: 1350  
Sequence: 1 atggacagagtttatgaaat.....caaggcaaaagtgtgtctga 1350

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmv:\*  
5: em\_estov:\*  
6: em\_estpi:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	301.8	22.4	1711	11	AK007734 Mus muscu
2	293	21.7	1774	11	BC027097 Mus muscu
3	259.4	19.2	624	10	BB200527 Mus muscu
4	245.4	18.2	810	12	EG675952 Mus muscu
5	240.6	17.8	676	9	AL652777 AL652777
6	236.8	17.5	392	10	BE031611 BE031611

7	198.8	14.7	881	13	BI827782
8	197.8	14.7	201	9	AA256009
c	193.4	14.3	678	9	AL638533
10	182.6	13.5	793	13	BI546216
11	173.2	12.8	385	9	AA791818
12	170.4	12.6	691	17	CNS03LFW
13	164.6	12.2	855	13	BI463670
c	163.2	8.4	691	17	AG085155
14	113.2	8.0	666	13	BI461092
15	107.6	7.9	1035	17	CNS01TL3
16	106	7.6	346	14	H87640
17	102	7.6	535	14	BM683941
c	88.6	6.6	537	14	BM931807
19	88.6	6.6	921	12	BF672674
20	77.6	5.7	921	12	BF672674
c	73.2	5.4	447	9	AI697050
21	73.2	5.4	447	9	AI697050
c	59.6	4.4	869	17	CNS032YG
23	56.8	4.2	387	17	AO087397
24	56.2	4.2	694	10	BB638337
25	53	3.9	627	12	BF211581
c	49.8	3.7	574	17	AZ593746
26	49.4	3.7	550	10	BB611251
c	48.6	3.6	857	17	CNS03K4E
28	46.8	3.5	755	9	AL653785
c	46.2	3.4	930	14	BQ718536
31	43.4	3.2	431	10	AW374923
c	43.4	3.2	464	9	AI923983
32	43.4	3.2	588	9	AI831578
c	43.4	3.2	588	12	BF195029
34	43.4	3.2	589	14	BM842656
c	43.4	3.2	609	10	AW966092
35	43.4	3.2	886	13	BI753850
c	43.4	3.2	917	9	AL551219
37	43.4	3.2	922	9	AL556942
c	43.4	3.2	926	12	BE735816
38	43.4	3.2	993	13	BM472443
c	43.4	3.2	1009	9	AL547473
41	43.4	3.2	1042	13	BM470447
c	43.4	3.2	1149	13	BM466219
43	43.4	3.2	877	13	BI463669
c	43.4	3.2			
45	43.2	3.2			

## ALIGNMENTS

RESULT 1	AK007734	AK007734	1711 bp	mRNA	linear	HTC 19-JAN-2002
LOCUS	AK007734	Mus musculus	10 day old male pancreas cDNA, RIKEN full-length			
DEFINITION	AK007734	enriched library, clone:1810038K19:McLeod syndrome gene homolog, full insert sequence.				
ACCESSION	AK007734	AK007734	GI:12841469			
VERSION	AK007734.1	HTC, CAP trapper.				
KEYWORDS	HTC, CAP trapper.	Mus musculus (strain:CS7BL/6J) 10 day old male pancreas cDNA to				
SOURCE	HTC, CAP trapper.	mRNA, clone lib:RIKEN full-length enriched mouse cDNA library				
ORGANISM	AK007734	Clone:1810038K19.				
REFERENCE	AK007734	Mus musculus				
AUTHORS	AK007734	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
TITLE	AK007734	Carninci, P. and Hayashizaki, Y.				
JOURNAL	AK007734	High-efficiency full-length cDNA cloning				
MEDLINE	AK007734	Meth. Enzymol. 303, 19-44 (1999)				
PUBMED	AK007734	99279253				
REFERENCE	AK007734	10349636				
AUTHORS	AK007734	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,				
TITLE	AK007734	Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
JOURNAL	AK007734	Normalization and subtraction of cap-trapper-selected cDNAs to				
MEDLINE	AK007734	prepare full-length cDNA libraries for rapid discovery of new genes				
PUBMED	AK007734	Genome Res. 10 (10), 1617-1630 (2000)				
	AK007734	20499374				
	AK007734	11042159				







```

QY 505 AAAGTATGTACAGATCCAGCCTTCTGGCTCAGTGGCCAGCTGACCTATCAGCTC 564
Db 181 AAAGTATGTACAGATCCAGCCTTCTGGCTCAGTGGCCAGCTGACCTATCAGCTA 240

QY 565 TATGTAGCCTGATCTCTGCAGAGGTTCCTCCCTGGGTAGAGTTGTG 609
Db 241 TATGTAGTCTGATCTCAGCAGAGTCCCTCCCTGGGTAGAGTTGTG 285

RESULT 4
BG675952 810 bp mRNA linear EST 01-MAY-2001
LOCUS 602623310F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4747323 5',
DEFINITION mRNA sequence.
ACCESSION BG675952
VERSION BG675952.1 GI:13907348
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 810)
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-i@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10597 row: c column: 04
High quality sequence start: 4
High quality sequence stop: 797.
Location/Qualifiers
source
1. .810
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:4747323"
/clone_lib="NCI_CGAP_Skn4"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (TI phage-resistant)"
/notes="Organs: skin; Vector: pCMV-Sport6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 214 a 178 c 169 g 249 t
ORIGIN
Query Match 18.2%; Score 245.4; DB 12; Length 810;
Best Local Similarity 99.2%; Pred. No. 4.2e-55;
Matches 257; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1092 CTTGGTTTTTAAGTCTTTGGAGTGAAGTTACTGAATTACTGTCTATCTTCCTTGATTC 1151
Db 1 CTTGGTTTTTAAGTCTTTGGAGTGAAGTTACTGAATTACTGTCTATCTTCCTTGATTC 60

QY 1152 CTTGCAGCTCATTTATGCTTATCTGATTTCCATTGATTCATGCTCCTTTTCTTCCAGTA 1211
Db 61 CTTGCAGCTCATTTATGCTTATCTGATTTCCATTGATTCATGCTCCTTTTCTTCCAGTA 120

QY 1212 CTTGCATCCATTGGGCTCACTCTTCACCCATATGTAGTAGACTACCTCCATTTGTCTG 1271
Db 121 CTTGCATCCATTGGGCTCACTCTTCACCCATATGTAGTAGACTACCTCCATTTGTCTG 180

QY 1272 CTTGTACACGACCCCTGGACCGGTTGAGACTCAGAGCCACCTTTGAGACTGAAGC 1331
Db 181 CTTGTACACGACCCCTGGACCA -GGTTGAGACTCAGAGCCACCTTTGAGACTGAAGC 239

```

```

QY 1332 AAGGCAAGTGTGTCTGA 1350
Db 240 AAGGCAAGTGTGTCTGA 258

RESULT 5
AL652777 676 bp mRNA linear EST 13-DEC-2001
LOCUS AL652777 XGC-gastrula Silurana tropicalis cDNA clone Tgas029g15 5',
DEFINITION mRNA sequence.
ACCESSION AL652777
VERSION AL652777.1 GI:17663442
KEYWORDS EST.
SOURCE western clawed frog.
ORGANISM Silurana tropicalis
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
TITLE 1 (bases 1 to 676)
JOURNAL Xenopodinae; Silurana.
COMMENT 1. (bases 1 to 676)
Huckle, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.
Sanger Xenopus tropicalis EST project 2001 (10_2001)
Unpublished (2001)
Contact: Huckle E
Sanger Centre
Hinnton, Cambridgeshire, CB10 1SA, UK
Email: tropesanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: Tgas029g15.sp6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
source
1. .676
/organism="Silurana tropicalis"
/db xref="taxon:8364"
/clone="Tgas029g15"
/clone_lib="XGC-gastrula"
/dev_stage="gastrula (stages 10.5-13 mixed)"
/lab_host="Escherichia coli XL1-blue"
/notes="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dT primed from Sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."

BASE COUNT 138 a 159 c 155 g 224 t
ORIGIN
Query Match 17.8%; Score 240.6; DB 9; Length 676;
Best Local Similarity 59.9%; Pred. No. 7.8e-54;
Matches 402; Conservative 0; Mismatches 269; Indels 0; Gaps 0;

QY 604 GTTGTGTAATGTTATTTTCCTGGTATCTGTCACTATGGGCGCACCTTTGCAATATG 663
Db 6 GGTGTCTCATGGTATGTGTCTTGCCTCAACTACATATGGTCCCTTAATCCTAATATGC 65

QY 664 TTGCTATCCAGATCAAGTAGTACGACTACAAAGATTGCCCTTGGGCCACTAGAAAGTCTC 723
Db 66 CTGGCCATACAAATTAAGTAGTAGTACTATAAAGTTCGTCTAAGTGGCCAGCGTTCTG 125

QY 724 TGCATACCATCTGGCGGACATTGGAGATCACTTTCCGCCCTCCCTGATTCGGTGCCTTC 783
Db 126 TGCATCGTAGTGGAGATGTCTGGAGATTGCTACCCGTTTACAGATTCTGGTCTCTTTT 185

QY 784 TCAGCCACTTTGAATTTGAAGGCTGTGCCCTTCTCTAGTGTCTCACTCTCTCATCTC 843
Db 186 TGTTAGCTCTTAAAGCCTGGGTGGCTGCAGTGGCCCTTAGCAAACTTACTCGCGCTCTC 245

QY 844 TTTGAGCCCTTGAATTAAGTTCTGGAGAAAGTGGTCCAGATGCCCAATAAATAGAGAA 903
Db 246 CTTTGTCTTGGGTTAAATTTCTGAGCAGTGGGACAGCCTGCCCCCAAAACATAGAAAG 305

QY 904 AACTTCAGCCGGGTGGGCACTCTGGTGGTCTCTGATTTTCAGTCACCATCTCTATGCTGC 963

```

```

Db 306 AACTTTAGCTGGTTGGTACTGTGTACAGTCTCTTGGGTGAGTAACCTTGCTTTATTACGCC 365
QY 964 ATCAACTTCTCTTGTCTGTCTAGCTTTTGAGTTTGAGTTGGCAGACAGAGATCTCTGTCGAC 1023
Db 366 ATCAATATGTTTCTGTCTGTCTCTTCTGCTTACAGTAAACCTATCAGATCGGAGCTTAATTGAC 425
QY 1024 AAGGGCAGAACTGGGACATATGGGCCCTGCACTATAGTGTGAGGTTGGTAGAGAAATGTG 1083
Db 426 CAGTCACAAAACCTGGGAGGTTAAGTCTTCAATTATACCAATTCGGTGTCTCGAGAATGCA 485
QY 1084 ATCATGCTCTGGTTTAAAGTCTTTGGAGTGAAGTGTCTACTGAATTAATCTGTCATTCCTC 1143
Db 486 GCCCTCATCTGCTCTGTGTACTCTACAAGGAGGATGTGTTTGAGTATTTCTTGTTCCCA 545
QY 1144 TTGATTGCTTGGCAGCTCATTTATCTTATCTGATTTTCCATTGACTTCTATGCTCTCTTTTC 1203
Db 546 CTGCTGGTGTCAACTCTTAGTAGATATGCCCGGCACTCTTCTCATGCTGCTTTT 605
QY 1204 TTCAGTACTTGCATCCATTCGCTCACTCTTCCACCCATAATGTAGTAGACTTACCTCCAT 1263
Db 606 TACCAGTACCTTCAACCTTGGCGCTTCTCTTTCAGCCACAGGTTGAAGACTGTTGGCT 665
QY 1264 TGTCTGCTG 1274
Db 666 TGTGTTGCTG 676

RESULT 6
LOCUS BE031611 392 bp mRNA linear EST 09-JUL-2000
DEFINITION 130228 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BE031611
VERSION BE031611.1 GI:8326620
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 392)
Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.,
and Keele, J.W.
TITLE Design and use of two pooled tissue normalized cDNA libraries for
JOURNAL EST discovery in swine
COMMENT Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACGATCAGGAC
Plate: 59 row: C column: 17
Seq primer: ATTTAGTGACACTATAG.
FEATURES
Location/Qualifiers
1..392
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1PIG"
/tissue_type="pooled"
/lab_host="DH10B"
/note="vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 81 a 87 c 100 g 124 t
ORIGIN
Query Match 17.5%; Score 236.8; DB 10; Length 392;

```

```

Best Local Similarity 90.4%; Pred. No. 6.7e-53;
Matches 253; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
QY 1071 GGTAGAGAAATGATCATCGTCTTGGTTTAAAGTTCTTTGGAGTGAAGTGTACTGAA 1130
Db 2 GGTAGAGAACGGTATCATGGTCTTGGTTTAAAGTTCTTTGGAGTGAAGTGTACTGGA 61
QY 1131 TTACTGTCAATCTCTTGGTTGCTTCAGCTCAATATTGCTTATCTGATTTCATTGACTT 1190
Db 62 TTACTGTCAATCTCTTGGTTGCTTCAGCTCAATATTGCTTATCTGATTTCATTGACTT 121
QY 1191 CATGCTCTTTTCTTCCAGTACTTGCATCCATTTGGCTCACTCTTCCACCATAATGTAGT 1250
Db 122 CATGCTCTTTTCTTCCAGTACTTGCACCGCTGCTCACTCTTCCACCACACGTAGT 181
QY 1251 AGACTACCTCATTTGCTGTCTGTGCACAGCACCTCGGACAGGTTTGAGAACTCAGA 1310
Db 182 GGACTACCTACACTGTGTCTGTGCATCAGCACCTCGGGCAGGGTTTGAGAACTCAGA 241
QY 1311 GCCACCTTTTGAGACTGAAGCAAGGCAAGTGTCTCTGA 1350
Db 242 GCCATCTGTTGATGCTGAACAAAGCAGACGATTTGCTGA 281

RESULT 7
LOCUS BI827782 881 bp mRNA linear EST 04-OCT-2001
DEFINITION 603075458F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5167070 5',
mRNA sequence.
ACCESSION BI827782
VERSION BI827782.1 GI:15939332
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 881)
NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1415 row: d column: 15
High quality sequence stop: 782.
FEATURES
Location/Qualifiers
1..881
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5167070"
/clone_lib="NIH_MGC_119"
/tissue_type="medulla"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: EcoRV (destroyed); RNA source normal medulla from
anonymous male age 27. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.3 kb, insert size range
0.9-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 013. Note:
this is a NIH_MGC Library."
BASE COUNT 257 a 163 c 194 g 266 t
ORIGIN
Query Match 14.7%; Score 198.8; DB 13; Length 881;
Best Local Similarity 57.3%; Pred. No. 1.7e-42;

```

Matches 444; Conservative 0; Mismatches 313; Indels 18; Gaps 4;

```
Qy 46 GTTTTCATCTCTGGAGGAGATGTCATCGTGGAGCCACCCCGATTTACTTTTCCATTT 105
Db 110 GTTTTCATCTCTGGAGGAGAAATAGTCTCTGG-CAGAGACTCCATTAAGTCTTCCATTT 168
Qy 106 AGCATCCTTTTCTCCACCTTTTGTACTGTGGGAGGCTGCATCTGCTTGTACATGTTT 165
Db 169 AGCATATCTCTCAACTGTTCTCTACTGTGTGAGTGGCTTGTGTATACATGTTT 228
Qy 166 AGAATCTATCGAAGAAATAGTGAACCTTACCGGATGACATACACTTTTCTTCTTTATG 225
Db 229 GAAATTTATCGAAGAAATAGTGAACCTTCTGATGTCTTACCATCAGCTTTATTTATT 288
Qy 226 TTTTTCATCCATTTATGTCAGCTTCACTTTTGTGTCACAGAGATAGCCAAAGAT 285
Db 289 GTGGGGCAATTTTGGATCAAAATATCTGATGTTTTTCAACAAGAACTTGAGGAGAAAT 348
Qy 286 AAACCGCTATCATTTATTTATGATCTAATCTCTTGGGACCTGTTATCAGATGTTTGGAG 345
Db 349 AAGGCTGCATTTACTTTTGTGGCAATTTCTTTTAGGACCTATNGTGAGGTGTTTGCAC 408
Qy 346 GCCATGATTAAGTACCTCACACTGTGGAAGAAAGAGGAGGAGGCCCTATGTGAGC 405
Db 409 ACCATTAGAAATTTACCAAAATGCTTGAATAATCTTAAACAGGAGAAAGAGACTCAA 468
Qy 406 CTCACCGAAGAGATGCTAATAGATGGCGAGAGGTGCTGATAGATGGGAGGTGGC 465
Db 469 GTTAGCATCAAAAGA-----GAAACACGATGCTGGAAGGGAGATTTGCA 513
Qy 466 CACTCATCCGACCCCTGGCTATGCAACCGCAATGCTTACAAACGTATGTACAGATCAA 525
Db 514 TTCTCAATCCGGATAAATTTCTGACAGAGAGGCTTTCAAGTACATGTCTAGTATTCAG 573
Qy 526 GCCTTCTCGGCTCAGTGCCTCCAGCTGACCTATCAGCTCTATGTGAGCCTGATCTCTGCA 585
Db 574 GCCTTCTCGGCTTCTGTTCCACAATTAATTTTGCAGATGTATATCAGTCTCACTATACA 633
Qy 586 GAGGTTCCCTCGGTAGAGTTGTCTAATGTTATTTCCCTGGTATCTGTGACCTATGGG 645
Db 634 GAATGGCCTTTGAATAGAG-CATGCTGATGACATGTTCCCTGTTTATCAGTTACTTGGG 692
Qy 646 GCCACCTTTTGCATATGTTGGCTATCCAGATCAAGTACGATGACTACAAGATTCGCTT 705
Db 693 GCATTCGCTGCAATATACGTGGCATCCAGATCAGCAATGATGATATCACCATAAGCT 752
Qy 706 GGCCCACTAGAAGTCTCTCT-GCATCACCATCTCTGGCGACATTTGGAGATCACTTCCCGCT 764
Db 753 ACCGCGATAGAATCTCTGCTGCTGCTGAGTGGGCGAGTTATGGAAGTTATCTCAGCTGT 812
Qy 765 CTGATTTCTGGTCTCTTCTCAGCACTTTGAAATTTGAAGCTGTGCTTCTCTTA 819
Db 813 AGTGACTCTGGCATTTTCCCTTGCATCTCTGCACTTGTAGAGCTTAACCGTTGGTA 867
```

RESULT 8  
AA256009  
LOCUS  
DEFINITION  
VERSION  
ACCESSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AA256009  
zs28c06.r1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:686506 5',  
mRNA sequence.  
AA256009  
AA256009.1 GI:1891549  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 201)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.

Email: [cgapsb-r@mail.nih.gov](mailto:cgapsb-r@mail.nih.gov)  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert length: 1241 Std Error: 0.00  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 183.  
Location/Qualifiers

FEATURES  
source

1..201  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:686506"  
/clone\_lib="NCI\_CGAP\_GCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD-),  
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
primed with a Not I - oligo(dT) primer  
[5'-TGTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTTTTTT-3',  
]. Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaudo."

BASE COUNT 39 a 46 c 36 g 80 t  
ORIGIN

Query Match 14.7%; Score 197.8; DB 9; Length 201;  
Best Local Similarity 99.0%; Pred. No. 1.8e-42;  
Matches 199; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1077 GAATGTGATCATGTCTTGTGTTTAAAGTTCTTTTGGAGTGAAGTGTACTGAATTAAGT 1136  
Db 1 GAATGTGATCATGTCTTGTGTTTAAAGTTCTATGAGTGAAGTGTACTGAATTAAGT 60  
Qy 1137 TCATTCCTGATTCGCTTGCAGCTCATTTGCTATCTGATTTCCATTTGATTCATGCT 1196  
Db 61 TCATTCCTGATTCGCTTGCAGCTCATTTGCTATCTGATTTCCATTTGCTGCTCATGCT 120  
Qy 1197 CCTTTTCTCCAGTACTTGCATCCATTCGCTCACTCTTACCCCATATGATAGACTA 1256  
Db 121 CCTTTTCTCCAGTACTTGCATCCATTCGCTCACTCTTACCCCATATGATAGACTA 180  
Qy 1257 CCTCCATTGTGCTGCTGCTCA 1277  
Db 181 CCTCCATTGTGCTGCTGCTCA 201

RESULT 9  
AL638533/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AL638533  
XGC-neurula Silurana tropicalis cDNA clone TNeu018n12 5',  
mRNA sequence.  
AL638533  
AL638533.1 GI:16790512  
EST.  
western clawed frog.  
Silurana tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
Xenopodinae; Silurana.  
1 (bases 1 to 678)  
Huckle, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.  
Sanger Xenopus tropicalis EST project 2001 (10\_2001)  
Unpublished (2001)  
Contact: Huckle E  
Sanger Centre  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: [trop@sanger.ac.uk](mailto:trop@sanger.ac.uk)  
Sanger xenopus tropicalis EST project 2001

```

TROPICALIS_SEQUENCE_ID: TNeu018n12.sp6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
  1..678
    /organism="Silurana tropicalis"
    /db_xref="taxon:8364"
    /clone="TNeu018n12"
    /dev_stage="neurula"
    /lab_host="Escherichia coli DH10B"
    /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dT primed from 5' of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."
BASE COUNT      225 a 156 c 153 g 144 t
ORIGIN
Query Match      14.3%; Score 193.4; DB 9; Length 678;
Best Local Similarity 58.7%; Pred. No. 4.3e-41;
Matches 335; Conservative 0; Mismatches 236; Indels 0; Gaps 0;
Qy 603 AGTTGTGCTAATGGTATTTCCCTGGTATCTGTCACTATGGGGCCACCCCTTTGCAATAT 662
Db 616 AGCTGTCTCATGGCTATGTGCTTGCCTCAACTACATATGTGCTTAACTCTTAATGT 557
Qy 663 GTTGGCTATCCAGATCAAGTACGATGACTACAAAGATTCGCTTGGGCCACTAGAAATCCT 722
Db 556 CTGGCCATACAAATTAAGTATGATGACTATAAAGTTCGTCTAAGTGTGCCAGCGTTCCT 497
Qy 723 CTGCATACCACTCTGGCGACATCGAGATCACTTCCCGCCCTCATCTCTGTGCTCTT 782
Db 496 GTGCATCGTATGTGGAGATGCTGGAGATGTCTACCCGTGTACAGTTCTGTGTTCTTT 437
Qy 783 CTCAGCCACTTTGAATTTGAAGCTGTGCGCTTCCCTAGTGTCTCAACTTCTGTATCATCT 842
Db 436 TTGTTCAAGCTTTAAGCCCTGGGTGCTGAGTGGCTTAGCAAACTTACTGGCGTCTT 377
Qy 843 CTTTGAGCCCTGGATTAAGTTCTGGAGAAAGTGGTGCCAGATGCCCAATAACATGAGAA 902
Db 376 CTTTGTGCTTGGTTAAATTTCTGGAGCAGTGGGACAGCTGCCCAAAACATAGAAAA 317
Qy 903 AAACCTTACGCGGGTGGCAGCTCTGTGTCTGCTGATTTAGTCACCATCTCTATGCTGG 962
Db 316 GAACCTTACGCTTGGTACTGTGACAGTCTCTTGGGTGAGTAACCTTGTCTTATTCAGC 257
Qy 963 CATCAACTCTCTGCTGTGCTTGCAGTTGAGTTGGCAGACAGAGATCTGCTCGA 1022
Db 256 CATCAATATGTTTGTGCTGTGCTTTCAGCTAAACCTATCAGATCGGGAATTAATTGA 197
Qy 1023 CAAAGGGCAGAACTGGGGACATATGGGCTCGCACTATAGTGTGAGGTTCGTTAGAGAAATGT 1082
Db 196 CCAGTCACAAACTGGGGAGGTAAAGTCTTCAATATACCAATCGGTTCGCGAGAAATGC 137
Qy 1083 GATCATGCTCTGTTGTTTAAAGTTCTTTGGAGTGAAGTGTCTACTGAATTAATCTGCTATTC 1142
Db 136 AGCCCTCATCTCTGCTGTGCTTCTACAAAGGAGGATGTGTTTGAATTTCTGTTTCCCC 77
Qy 1143 CTTGATTGCTTGGCAGTCAATTAATGCTTAT 1173
Db 76 ACTGCTGTGTTCAACTCTTAGTAGGATAT 46

RESULT 10
LOCUS      B1546216
DEFINITION 60318602F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5259961 5',
            mRNA sequence.
ACCESSION  B1546216
VERSION    B1546216.1 GI:15433528
KEYWORDS   EST.
SOURCE     human.

```

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 793)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Plate: LLAM11655 row: c column: 02  
High quality sequence stop: 707.

FEATURES  
Location/Qualifiers  
1..793  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5259961"  
/clone\_lib="NIH\_MGC\_95"  
/tissue\_type="hippocampus"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site 1: BamHI; Site 2: Sali-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.5 kb and normalized to 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH\_MGC Library."

BASE COUNT 180 a 215 c 188 g 210 t  
ORIGIN

Query Match 13.5%; Score 182.6; DB 13; Length 793;  
Best Local Similarity 56.9%; Pred. No. 3.8e-38;  
Matches 419; Conservative 0; Mismatches 304; Indels 14; Gaps 4;

Qy 250 ACCCTCATTTTGTCCACAGAGATCTAGCCAAAGATAAACCGCTATCATTTATTCAT 309  
Db 1 AGCGGCTTCTTCTGTACACCGCAGCTTCAGCCGCGACCGCCGCTCGTACTGCTGTCAC 60  
Qy 310 CTAATCTCTTGGGACCTGTATCAGATGTTTGGAGGCCATGATTAAAGTACCTCAGACTG 369  
Db 61 CTGCTGCAACTTGGGCGCCGTTTTCAGGTGTTTGAAGTCTTCTGCACTCTAC----- 111  
Qy 370 TGGAAAGAAAGAGGAGCAGGAGGAGCCCTATGTACAGCTTCACCCGCAAGAGAG--ATGCTA 426  
Db 112 TTTCAGTCAGGCAACAATCAAGAGCCTTATGTCAGTATCACCAAGAGAGGCAAAATGCCA 171  
Qy 427 ATAGATGGCGAGAGGTGCTGTATAGAAATGGAGGTGGGCACCTCCATCCGAGCCCTGGCT 486  
Db 172 AAAAAATGGCCCTCTCAGAGGAGATTGAGAAGGAGGTGGGCGCAGAGGCAAACTAATC 231  
Qy 487 ATGCACCGCAATGCCCTACAAAGCTATGTACAGATCCAGAGCTTCCTGGGCTCAGTGCCTC 546  
Db 232 ACCACCGATCAGGTTTACGCGGGGTCGGTGATCCAGGCTTCTTGGGCTCAGCCCC 291  
Qy 547 CAGCTGACCTATCAGCTCTATGTAGCCCTGATCTCTGACAGAGTTCCTCCCTGGGTAGAGTT 606  
Db 292 CAGCTGACCTACAGCTGTACATAAGTGTATGTCAGCAGGAGCTCACTGTTGGAGAAGT 351  
Qy 607 GTGCTAATGGTATTTCCCTGGTATCTGTCACTATGGGCCACCCCTTTCGCAATATGTTG 666  
Db 352 CTCCTCATGACCATATCCCTGTTGTCCATTGTGTAGGAGCTTGGGCTGCAACATCCTA 411  
Qy 667 GCTATCCAGATCAAGTACGATGACTAC--AGATTCCCTTGGGCGCACTAGAAAGTCTCTG 725



```

Db 412 GCCATCAAAATCAAGTACGATGATGAAGTCAAGAGTGAAGCCCTCTGGCCTATGCTGT 471
Qy 726 CATACCATCTGGCGACATTCGAGATCACTTCCCGCCTCCGATTCGTGCTCTCTTC 785
Db 472 TATCTTCTGTGGAGAGCTTTGAGATTGCCACTGCACTGAGTTGAGTCTGCTCTTTTAC 531
Qy 786 AGCCATTTGAAATTGAAGGCTGTGCCCTTCCTCTAGTCTCAACTTCTCTGATCATCTCTT 845
Db 532 CTCCTCTCTGAAGACCTGGCGGTGGTTATATACTCATCAACTTCTTCAGTTTCTCTTA 591
Qy 846 TGAGCCCTGGATTAGTTCTGAGAGAGTGGTGGCCAGATGCCAATAACATTGAGAAAAA 905
Db 592 GTACCCCTGGATCCCTCTTCTGTGTCAGTGGTTCCCCATTCCCTGAGAACATAGAG-AA 650
Qy 906 CTTAGCCGGGTGGCAGCTCTGCTGCTGCTGATTTTCAGTCACCATCTCTATGCTGGCAT 965
Db 651 CCTAGTAGAGTGGCAGCACCACCATTTGACTATGCTTAACTTTACTTCTACTCTGGTA 710
Qy 966 CAACTTCTCTTCTGCTGT 982
Db 711 TTAACCATGTTCTCTGCT 727

RESULT 11
AA791818 385 bp mRNA linear EST 09-FEB-1998
LOCUS v61c09.r1 Stratagene mouse skin (#937313) Mus musculus CDNA clone
DEFINITION IMAGE:1150768 5' similar to SW:HK_HUMAN P51811 MEMBRANE TRANSPORT
PROTEIN XK ; mRNA sequence.
ACCESSION AA791818
VERSION AA791818.1 GI:2854773
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 385)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Seftoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:623976
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 378.
FEATURES
source
1. .385
/organism="Mus musculus"
/strain="C57BL/6"
/db xref="taxon:10090"
/clone="IMAGE:1150768"
/clone_lib="Stratagene mouse skin (#937313)"
/sex="females"
/tissue_types="whole skin"
/dev stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/notes="Organ: skin; Vector: pBluescript SK-; Site:1: EcoRI
; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo
dT. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"

```

```

BASE COUNT 106 a 82 c 100 g 97 t
ORIGIN
Query Match 12.8%; Score 173.2; DB 9; Length 385;
Best Local Similarity 93.3%; Pred. No. 1e-35;
Matches 181; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Qy 416 AGAAGATGCTAATAGATGGCGAGGAGTCTGATAGAATGGGAGGTGGCCACTCCATCC 475
Db 1 AGAAGATGCTAATAGTGTGCCAGGAGGTCTGATAGAATGAAGGTTGGGCCACTCCATCC 60
Qy 476 GGACCTCTGCTATGACCGCAATGCTTCAAAACGATGTGTACAGATCCCAAGCCTTCTCTGG 535
Db 61 GGACCTCTGCTATGCTCGCAATGCTTCAAAACGATGTGTACAGATCCCAAGCCTTCTCTGG 120
Qy 536 GCTCAGTCCCCAGCTGACCTATCAGCTCTATGTGAGCCTGTCTCTGAGAGTTTCCCC 595
Db 121 GCTCAGTCCCCAGCTGACCTATCAGCTATATGTGATCTGTGATCTCTCAGCAGAGTCCCC 180
Qy 596 TGGGTAGAGTTGTTG 609
Db 181 TGGGTAGAGTTGAG 194

RESULT 12
CNS02LFFV 691 bp DNA linear GSS 14-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 147116 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL202756
VERSION AL202756.1 GI:7861101
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 691)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 691)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 691)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000)
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.
FEATURES
source
1. .691
/organism="Tetraodon nigroviridis"
/db xref="taxon:99883"
/clone="147116"
/clone_lib="G"
/notes="Genoscope sequence ID : COAG147BE08LP1-end : T7"
BASE COUNT 128 a 182 c 187 g 189 t 5 Others
ORIGIN
Query Match 12.6%; Score 170.4; DB 17; Length 691;
Best Local Similarity 58.9%; Pred. No. 7.1e-35;

```

Matches	305;	Conservative	2;	Mismatches	210;	Indels	1;	Gaps	1;
Qy	590	TTCCCTCGGGTAGAGTCTTGCTAAATGGTATTTTCCCTGTATCTCTCACCTATGGGGCCA	649						
Db	138	TGCTTCGTGTTGCGAGTGCTCTGATGATCATTTAACTTGCTCCCATCACATACGGGGCTC	197						
Qy	650	CCCTTTGCAATATGTTTGGCTATCCAGATCAAGTAGCATCAAGAATTCGCCTTTGGGC	709						
Db	198	TGGTGTGCRAGCTCTCGGCCATCCACATCAGATACGACGACTTACAAGGTGGG-GGTCCGC	256						
Qy	710	CACTAGAAGTCTCTGTCATCACCATCTCGCGGACATTCGGAGATCACTTCCCGCCCTCTGA	769						
Db	257	CCCGCGCCTACTGTGTCATGACCGGTGTGGAGAGGTTGGAGATGCCACAGGATCAGGG	316						
Qy	770	TTCTGGTGCTCTTCTCAGCGCACTTTGAAATGGAAGGCTGTGCCCTTCTCTAGTGTCTCAACT	829						
Db	317	TTCTGGTCTCTTCAGCAGGGGCTCAATACTGGGTGNCCTGCTGGGTCCGNCACCC	376						
Qy	830	TCCTGATCATCCTCTTTGAGCCCTGGATTAAGTTCTCGAGAGAAGTGGTGCCAGATGCCCCA	889						
Db	377	TGCTCTCTTCTTCTTCTTCCAAACCTCGGACCGAGTTCTGGGCCAAGAAAGGCTCGTTGACTC	436						
Qy	890	ATAACATTGAGAAAAAATTTCAGCCGGGTGCGGACATCTCGTGGTCTGTATTCAGTCACCA	949						
Db	437	AGGGCCTGGAGCGCAAACTTCTCTCATTTTCGGCACCGTGTGGTGTGAGCCTGTTTCAGC	496						
Qy	950	TCCTCTATGCTGGATCAACTTCTCTGCTGTCAGCTTTCAGTGTGAGGTTGGCAGACA	1009						
Db	497	TGCTGTTTCGCTCGCTGCMTCAACGTGTTCTCTGCTGCGCGGTGAGCTGAGCTTCACCCACC	556						
Qy	1010	GAGATCTCGTCCGACAAAGGGCAGAACTCGGGACATATGGGGCTGCACATATAGTGTGAGGT	1069						
Db	557	GGGAGCTGATCCAGAGGAGGAGGGCTGGGGCCGACTGGCTCTGTACTTCTGCGCAGCCT	616						
Qy	1070	TGGTAGAAGATGTGATCATGTGCTTTGGTTTTTAAGTTC	1107						
Db	617	TCGTGGAGAACTTTGTCTCTCATCAGCCTGTGGTACTTC	654						

RESULT 13  
BI463670 855 bp mRNA linear EST 21-AUG-2001  
LOCUS 603207340F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:5273312 5',  
DEFINITION mRNA sequence.  
ACCESSION BI463670  
VERSION BI463670.1 GI:15254326  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 855)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.

```

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11689 row: 0 column: 09
High quality sequence stop: 826.
Location/Qualifiers
1. 855
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5273312"
/clone_lib="NH MGC 97"
FEATURES
source

```

```

/lab_host="DH10B"
/notes="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI [gtcagag
]; Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NIHGRI, National
Institutes of Health). Note: this is a NIH MGC Library."
257 a 156 c 186 q 256 t
BASE COUNT

```

Query Match	12.2%	Score 164.6	DB 13	Length 855	
Best Local Similarity	57.0%	Pred. No. 2.8e-33			
Matches 329	Conservative	0	Mismatches 234	Indels 14	Gaps 1
Qy	1	ATGGACAGAGTTTATGAAATTCCTGAGGAGCCAAATGTGGATCGGTTTCCATCTCTGGAG	60		
Db	104	ATGGAGACAGTGTGTTGAAGAGATGGATGAAGAAAGCAGGAGAGTTTTCATCTCGAAA	163		
Qy	61	GAAGATGTCATCCGTGGAGCCAACCCCGATTTATCTTTTCCATTTAGCATCCTTTTCTCC	120		
Db	164	GAAGAAATAGTCTCTGGCCAGAGACTCCATCTAAGCTTCTCTTTTAGCATTATCTTCTCA	223		
Qy	121	ACCTTTTGTACTGTGGGAGGCTGCATCTGCTTTGTACATGTTAGAAATCTATCGAAAG	180		
Db	224	ACTGTTCTCTACTGTGGTGAGGTGGCTTTGGTTTATACATGTTTGAATTTATCGAAAA	283		
Qy	181	AATAGTGAACACTTACCGGATGACATACACCTTTTCTTTCTTATGTTTTCATCAATATG	240		
Db	284	GCATAGACACATCTGGATGTCAATTACCATCAGCTTTATTATGTGGGGCAATTTTG	343		
Qy	241	GTCCAGTTGACCTTCATTTTGTGCCAGAGATCTAGCCAAAGATAAACCGCTATCATTA	300		
Db	344	GATCAAAATTAATCTGTATGTTTTCACAAAGACTGTGAGGAGAAATAGCGTGCATTACTT	403		
Qy	301	TTTATGCATCTAATCTCTTTGGGACCTGTTTATCAGATGTTTGAGGCCATGATTAGTAC	360		
Db	404	TTTTGGGACATCTCTTTTGTAGACCTATTGTGAGGTGTTTGACACCATTAGAAATTAC	463		
Qy	361	CTCACCTGTGGAAAGAGGAGGAGGAGGCCCTATGTGAGCCTCACCCGAAAGAAG	420		
Db	464	CACAAATGGTTGAAAAATCTTAAACAGGAGAGGAAGA-----GACTCAAG	509		
Qy	421	ATGCTAATAGATGGCGAGGAGTGTGATAGATGGGAGGTGGGCCACATCCATCCGGACC	480		
Db	510	TTAGCATCAAAAGAGAAACACGATGCTGGCAAGGGAGATTGCATTCCTCAATCCGGGAT	569		
Qy	481	CTGGCTATGCACCGCAATGCTTACAAACGTATGTCAAGATCCAAAGCTTCCTCGGGCTCA	540		
Db	570	AATTTTCATGACGACAGAGGCTTTCAAGTACATGTCAGTGATTCAGGCTTTCTCGGTTCT	629		
Qy	541	GTGCCCCAGCTGACCTATACAGTCTCTATGTGAGCCTGA	577		
Db	630	GTTCACAAATTAATTTTGCAGATGTGTATACAGTCTCA	666		

RESULT 14	LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
AG085155/c	AG085155	Pan troglodytes DNA, clone: PTB-083B22.F, genomic survey sequence.	691 bp	DNA	linear	GSS 03-NOV-2001		
	AG085155	Pan troglodytes DNA, clone: PTB-083B22.F, genomic survey sequence.						
	AG085155.1	GI:16636957						
	GSS.							
		Pan troglodytes male lymphoblast DNA, clone lib:PTB Chimpanzee Male						
		BAC Library clone:PTB-083B22.F.						
		Pan troglodytes						
		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;						
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.						
	1							
		Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,						
		Totoki, Y., Watanabe, H. and Sakaki, Y.						

TITLE BAC end sequences of Library PTB  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 691)  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.  
TITLE Direct Submission  
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: chimpbes@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
COMMENT Clones are derived from the chimpanzee BAC library PTB. This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.  
PRIMERS

Sequencing: -21M13

LIBRARY  
Vector : pKS145  
R.Site 1 : SacI  
R.Site 2 : SacI.  
FEATURES  
source  
1. .691  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-083B22.F"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC Library"  
BASE COUNT 230 a 151 c 127 g 181 t 2 others  
ORIGIN

Query Match 8.4%; Score 113.2; DB 17; Length 691;  
Best Local Similarity 60.9%; Pred. No. 2e-19;  
Matches 184; Conservative 0; Mismatches 118; Indels 0; Gaps 0;  
QY 601 AGAGTTGTGTAATGGTATTTCCCTGGTATCTGTGACCTATGGGGCCACCTTTGCAAT 660  
DB 354 ATAGCATTTGCTGATGACATGTTCCCTGTTATCAGTTACTTATGGGCCATTCGCTGCAAT 295  
QY 661 ATGTTGGCTATCCAGATCAAGTACGATGACTACAAGATTCGCCCTTGGGCCACTAGAATC 720  
DB 294 ATACTGGCCATCCAGATCAGCAATGATGATGATCACTACCACTAAGCTACCGCGGATGAATTC 235  
QY 721 CTCTGCATCACCATCTCGCGGACATTTGAGATCACTTCCCGCCTCTGATTCGTGTCGTC 780  
DB 234 TTCTGTGCTGATGTCGGCTTTTGGAGGTATCTCAGCTGAGTACTCTGGCATTT 175  
QY 781 TTCTGACCCACTTTGAATTTGAAGGCTGTGCCCCTTCTCTAGTGTCTCAACTTCCCTGATCATC 840  
DB 174 TTCATTGCATCTCTGAACTGAAGAGCTACCTCTTTTGTATCATATATTTTGTGCA 115  
QY 841 CTCTTTGAGCCCTGGATTAAGTTCTGAGAAAGTGTGCGCCAGATGCCCAATAACATTGAG 900  
DB 114 TTGTTGACACCGTGGCTGGAGTTTGGAAAAGTGAGGCTCGCGGAGCTGCTAGCTGCGG 55  
QY 901 AA 902  
DB 54 AA 53

RESULT 15  
BI461092  
LOCUS 603206970F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:5272840 5',  
DEFINITION mRNA sequence.  
ACCESSION BI461092  
VERSION BI461092.1 GI:15251748  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 666)

AUTHORS  
TITLE  
JOURNAL  
COMMENT

NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-romail.nih.gov  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM11688 row: k column: 17  
High quality sequence stop: 663.

FEATURES  
source

1. .666  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="INAGR:5272840"  
/clone\_lib="NIH\_MGC\_97"  
/lab\_hosts="DH10B"  
/notes="Organ: testis; Vector: pBluescriptR (modified  
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag  
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
size-selected for average insert size 2.2 kb and  
normalized to ROI 5. This is a primary library enriched  
for full-length clones and constructed using the  
Cap-trapper method (Carninci, in preparation). Library  
constructed by M. Brownstein (NIH/NHGRI, National  
Institutes of Health). Note: this is a NIH\_MGC Library."  
BASE COUNT 125 a 212 c 197 g 132 t  
ORIGIN

Query Match 8.0%; Score 107.6; DB 13; Length 666;  
Best Local Similarity 53.5%; Pred. No. 6.3e-18;  
Matches 277; Conservative 0; Mismatches 229; Indels 12; Gaps 2;  
QY 118 TCCACCTTTTGTACTGTGGGAGGCTGCATCTGCTTTGTATCATGTTAGATCATATCA 177  
DB 156 TCCGTGTTCTGTTGTTGGCCGAGACAACGGCGCGCTCAGCCTGAGCAGCACCTACCGC 215  
QY 178 AAGAAATAGTGAACCTTACCGATGACATACACCTTTCTTTCTTTATGTTTTCATCCATT 237  
DB 216 TCGGCGGGACCCGATGTGGCAGCGCTGACGTGTTTCTCGCTACTGCTTGCCTG 275  
QY 238 ATGCTCCAGTTGACCTCATTTTGTCCACAGAGATCTAGCCAAAGATAAAACCGCTATCA 297  
DB 276 CTGTCGAGCTCAGCTTCTCTTGTACACCGCAGCTCAGCCGCGCCGCTCGTA 335  
QY 298 TTAATTTATGATCTAATCTCTTTGGGACCTGTTATCAGATGTTTGGAGGCGCATGATTAG 357  
DB 336 CTGCTGTCGACCTGCTGCAACTTGGGCCCTTTTCAGGTGTTTGAAGTCTTCTGTCATC 395  
QY 358 TACTCACATGTTGGAGAAGAGGAGCAGGAGGAGCCCTATGTTCAGCCTCACCGGAAG 417  
DB 396 TACTTTC-----AGTCAGGCAACAATGAAGAGCTTATGTGCTATCACCAAGAAG 446  
QY 418 A---AGATGCTAATAGATGGGAGGCTGCTAGATAATGGGAGGTGGGCGCACTCCCATC 474  
DB 447 AGCAATATGCAAAATATGCTCTCAGAGAGATTTGAGAGAGGTGGGCCAGGAGAA 506  
QY 475 CGGACCTGCTGATGACCCCAATGCTTACAACTGATGTACAGATGCCAGCTTCTG 534  
DB 507 GGCAAACTAATACCCACCGATCAGCGTTTTCAGCCGGCGCTCGGTGATCCAGGCTTCTT 566  
QY 535 GCCTCAGTGGCCGAGCTGACCTATCAGCTTATGTGAGCCTGATCTCTGACAGAGTTCCC 594  
DB 567 GGCTCAGCCGCCAGCTGACCTTACAGCTGTACATAAGTGTGTCATGTCAGCAGGAGTCACT 626  
QY 595 CTGGGTAGAGTTGCTGCTTAATGTTATTTTCCCTGGTATC 632  
DB 627 GTTGGAGAAGAGTCTCTCTCATGACCATATCCCTGTTGTC 664

Search completed: March 30, 2003, 05:49:05  
Job time : 1638.01 secs